



PTO/SB/21 (09-04)

**TRANSMITTAL
FORM**

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	10/769,030
Filing Date	January 30, 2004
First Named Inventor	Oshima, Satoshi
Art Unit	2186
Examiner Name	Unassigned
Attorney Docket Number	16869S-104700US

ENCLOSURES (Check all that apply)

- | | | |
|---|---|---|
| <input checked="" type="checkbox"/> Fee Transmittal Form
<input type="checkbox"/> Fee Attached
<input type="checkbox"/> Amendment/Reply
<input type="checkbox"/> After Final
<input type="checkbox"/> Affidavits/declaration(s)
<input type="checkbox"/> Extension of Time Request
<input type="checkbox"/> Express Abandonment Request
<input type="checkbox"/> Information Disclosure Statement

<input type="checkbox"/> Certified Copy of Priority Document(s)
<input type="checkbox"/> Reply to Missing Parts/ Incomplete Application
<input type="checkbox"/> Reply to Missing Parts under 37 CFR 1.52 or 1.53 | <input type="checkbox"/> Drawing(s)
<input type="checkbox"/> Licensing-related Papers
<input type="checkbox"/> Petition
<input type="checkbox"/> Petition to Convert to a Provisional Application
<input type="checkbox"/> Power of Attorney, Revocation
Change of Correspondence Address
<input type="checkbox"/> Terminal Disclaimer
<input type="checkbox"/> Request for Refund
<input type="checkbox"/> CD, Number of CD(s) _____
<input type="checkbox"/> Landscape Table on CD | <input type="checkbox"/> After Allowance Communication to TC
<input type="checkbox"/> Appeal Communication to Board of Appeals and Interferences
<input type="checkbox"/> Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)
<input type="checkbox"/> Proprietary Information
<input type="checkbox"/> Status Letter
<input checked="" type="checkbox"/> Other Enclosure(s) (please identify below):
Renewed Petition to Make Special for New Application
Exhibit A, Search Report, Exhibit B, Table of Contents, and 7 Refs.
Return Postcard |
|---|---|---|

Remarks	The Commissioner is authorized to charge any additional fees to Deposit Account 20-1430.
---------	--

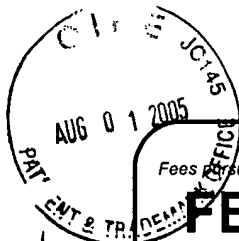
SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Townsend and Townsend and Crew LLP		
Signature			
Printed name	George B. F. Yee		
Date	July 28, 2005	Reg. No.	37,478

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature			
Typed or printed name	Cynthia McKinley	Date	July 28, 2005



Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL

For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$ 130

Complete if Known

Application Number	10/769,030
Filing Date	January 30, 2004
First Named Inventor	Oshima, Satoshi
Examiner Name	Unassigned
Art Unit	2186
Attorney Docket No.	16869S-104700US

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____
☒ Deposit Account Deposit Account Number: 20-1430 Deposit Account Name: Townsend and Townsend and Crew LLP

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☐ Charge fee(s) indicated below, except for the filing fee
☒ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17 ☒ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES Small Entity		SEARCH FEES Small Entity		EXAMINATION FEES Small Entity		Fees Paid (\$)
	Fee (\$)	Fee (\$)	Fee (\$)	Fee (\$)	Fee (\$)	Fee (\$)	
Utility	300	150	500	250	200	100	
Design	200	100	100	50	130	65	
Plant	200	100	300	150	160	80	
Reissue	300	150	500	250	600	300	
Provisional	200	100	0	0	0	0	

2. EXCESS CLAIM FEES

Fee Description	Small Entity	
	Fee (\$)	Fee (\$)
Each claim over 20 or, for Reissues, each claim over 20 and more than in the original patent	50	25
Each independent claim over 3 or, for Reissues, each independent claim more than in the original patent	200	100
Multiple dependent claims	360	180

Total Claims	Extra Claims	Fee (\$)	Fee Paid (\$)	Multiple Dependent Claims	Fee (\$)	Fee Paid (\$)
_____ -20 or HP = _____	X _____	= _____		_____	_____	_____

HP = highest number of total claims paid for, if greater than 20

Indep. Claims	Extra Claims	Fee (\$)	Fee Paid (\$)
_____ -3 or HP = _____	X _____	= _____	

HP = highest number of independent claims paid for, if greater than 3

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper, the application size fee due is \$250 (\$125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	Extra Sheets	Number of each additional 50 or fraction thereof	Fee (\$)	Fee Paid (\$)
_____ - 100 = _____	/ 50 = _____	(round up to a whole number) x _____	= _____	

4. OTHER FEE(S)

Non-English Specification, \$130 fee (no small entity discount)

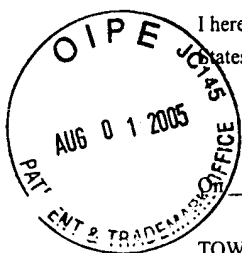
Other: Petition Fee

Fees Paid (\$)

130

SUBMITTED BY

Signature		Registration No. (Attorney/Agent) 37,478	Telephone 650-326-2400
Name (Print/Type)	George B. F. Yee		Date July 28, 2005



I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to:

PATENT
Attorney Docket No.: 16869S-104700US
Client Ref. No.: W1364-01EM

7/28/05
TOWNSEND and TOWNSEND and CREW LLP

By:

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Satoshi Oshima et al.

Application No.: 10/769,030

Filed: January 30, 2004

For: Storage Device Having a Capability
of Learning Access Patterns

Customer No.: 20350

Confirmation No. 7614

Examiner: unassigned

Technology Center/Art Unit: 2186

RENEWED
PETITION TO MAKE SPECIAL FOR
NEW APPLICATION PURSUANT TO
37 C.F.R. § 1.102(d) &
M.P.E.P. § 708.02, Item VIII,
ACCELERATED EXAMINATION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

A petition to make special the above-identified application in accordance with MPEP § 708.02, Item VIII, accelerated examination was filed May 13, 2005. A decision mailed June 24, 2005 denied said petition.

Submitted herewith is a renewed petition to make special the above-identified application in accordance with MPEP § 708.02, Item VIII, accelerated examination. The application has not received any examination by the Examiner.

(A) The Commissioner is authorized to charge the petition fee of \$130 under 37 C.F.R. § 1.17(h), and any additional fees that may be associated with this petition may be charged to Deposit Account No. 20-1430.

08/02/2005 EFLORES 00000122 201430 10769030

01 FC:1464 130.00 DA

(B) All the claims are believed to be directed to a single invention. If the examiner determines that all the claims presented are not obviously directed to a single invention, then Applicant will make an election without traverse as a prerequisite to the grant of special status where the specific grouping of claims will be determined by the examiner.

(C) A pre-examination search was performed by an independent patent search firm. A copy of the search report is provided herewith as Exhibit A. The pre-examination search includes a classification search, a computer database search, and a keyword search. The searches were performed on or around April 15, 2005, and were conducted by a professional search firm, Kramer & Amado, P.C. The field of search was directed to Class 711, subclasses 100, 101, 111, 112, 113 and 114 (U.S. & Foreign). Additionally, a computer database search was conducted on the U.S.P.T.O. systems EAST and WEST for U.S. and foreign patents; a keyword search was conducted in Class 707, and a literature search was also conducted on the internet and commercial databases for relevant non-patent documents. The field of search was confirmed by Examiner Stephen Elmer in Group Art Unit 2186. The following references were identified in the search report:

- (1) U.S. Patent Nos.:
 - 6,834,289 Kaneda et al.
 - 6,763,442 Arakawa et al.
- (2) U.S. Patent Application Publication Nos.:
 - 2003/0084237 Yoshida et al.
 - 2003/0120674 Morita et al.
 - 2003/0177330 Idei et al.
 - 2004/0193803 Mogi et al.
- (3) Literature
 - 3ware TwinStor™ Architecture by 3ware

(D) The above references are enclosed herewith, collectively as Exhibit B.

(E) Set forth below is a detailed discussion of the references, pointing out with particularity how the claimed subject matter recited in the claims, amended according to the preliminary amendment filed herewith, is distinguishable over the references.

Claimed Subject Matter of the Present Invention

The independent claims are 1, 10, and 13.

With reference to the disclosure, a storage device records data readout locations (i.e. a history) for a disk device to thereby learn access patterns. FIG. 1 illustrates a system having a storage device 101, a management computer 108, a network server 110, and a plurality of computers 111 connected through a network 109. See pg. 6, ln. 23 of the specification. The storage device 101 obtains an access history, such as data readout, from the storage device 101 for each computer 111 and then stores the access history for each computer in the cache control unit 106. See pg. 7 ln. 23-27. FIG. 2 illustrates a configuration of a history information saving table which saves information of the access history to the storage. The table includes a history information management table 301 and a history information saving list 302. See pg. 8, ln. 23-27.

With reference to the claims, a storage device records data readout locations (i.e. a history) for a disk device to thereby learn access patterns. Data corresponding to readout locations of the disk device are pre-read and stored in cache memory.

Claim 1 sets forth a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. Data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

Claim 10 sets forth that a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. When the computer reads out data from the storage device, the storage location of that data serves as history information. The maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The storage device reads out data to a cache memory based on the second command.

Claim 13 sets forth a read-ahead method in a storage device. A command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history. In the storage device, a location where data is to be stored is recorded as a history when the data is read out by the computer. A second command is

transmitted, specifying the computer and the history. In the storage device data is pre-read from the storage device, based on the computer and the history specified in the second command.

Detailed Discussion of References

U.S. Patent No. 6,834,289 to Kaneda et al.

Kaneda et al. relates to an information processing system and storage area allocating method comprising a plurality of storage devices and a library. FIG. 1 illustrates a system comprising three computers 101, 102, 151, three disk array systems 201, 202, 203, a library system 301 and a fiber channel switch 501. See col. 3, ln. 10-13 and col. 12, ln. 34-43. A cache disk 90 is also provided in computers 101, 102. FIG. 2 illustrates a conversion table 811 stored in computer 151. See col. 5, ln. 51-26. In order to read the data recorded on the medium stored in the library system 301, the computer 151 operates to instruct the library system 301 to convey the medium on which the data to be read is recorded from the shelf to the drive 30. Then, the computer 151 waits for a report on termination of the conveyance from the library system 301 and then instructs the drive 30 to read out the data. See col. 4, ln. 30-44.

As understood, the control unit of Kaneda et al. does not record data readout locations as a history for each computer based on a command containing information for specifying the computer that uses the storage device. Thus, Kaneda et al. does not set forth a storage device that records data readout locations (i.e. a history) in a cache memory for a disk device to thereby learn access patterns, as claimed.

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as

history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

U.S. Patent No. 6,763,442 to Arakawa et al.

Arakawa et al. relates to a data reallocation system used with multiple data storage systems such as RAIDs. FIG. 1 illustrates a computer and storage system comprising a host 100, disk arrays 200-1, 200-2, a switch 500, clients 800-1, 9800-2 and a local disk 190. Host 100 has a network interface 170 and a Fiber Channel interface 160 and is connected via network 700 to the clients 800-1, 800-2 and disk arrays 200-1, 200-2. See col. 5, ln. 42-54. The local disk 190 stores management information including a logical unit (LU) position name table 191 and an intra-LU address logical position name table 195 to be used by operating system (OS) 120 and file system 110. See col. 6, ln. 8-17. The manager 130 issues a write command of the SCSI standard to the LU for a logical volume and a row of a disk array via the FC interface 160 and writes parameters for information collection as data. See col. 9, ln. 18-22. When the disk array 200-2 receives a read command from the host 100 via FC interface 260-2, the control section 230-2 perceives that the read command concerns the LU and transfers the information readied on the memory 320-2 to the host 100. Control section 300-2 then reports the completion of the read to the host 100. See col. 9, ln. 38-45.

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

U.S. Pub. No. 2003/0084237 to Yoshida et al.

Yoshida et al. relates to a disk array system controller having a host switch interface section, disk array controlling units, and a cache memory section. FIG. 1 illustrates the disk array controller 1 comprising a plurality of disk array controlling units 1-2 and a host switch interface section 30. See section [0032]. The host switch interface section 30 refers to a management table 31 therein in response to the request made by the host computer 50 to find an optimum route to the cache memory14. See section [0033]. The host switch interface section 30

is connected through a PATH050, PATH151 and PATHBK 52 with the plurality of channel interface sections of the disk array controlling units. A management table 31 is provided in the host switch interface section 30 in which contains a path selection table 32, a history information table 33, wherein a path selection signal 40 is output to the path selection table 32 on the basis of selected signal PATH 41. See section [0034].

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

U.S. Pub. No. 2003/0120674 to Morita et al.

Morita et al. relates to a control method for a disk array device and a disk controller wherein a write history management table records histories of conducted writing processes for confirmation of a writing process conducted to a block. FIG. 1 illustrates a disk array device 102, a disk controller 103 and a subgroup of hard disks 114. See section [0054]. A write history management table 113 includes management information for judging whether or not data has been written in the magnetic disk 108. See section [0056]. FIG. 2 illustrates that prepared history flags 301 exist in certain subgroups of hard disks 114 for the M data stripes 201. FIG. 4 illustrates a history flag 301 value of "0" which indicates there is no history of data writing in the respective data stripe, and a flag 301 of "1" which indicates there is a history of data writing in the data stripe, indicating data has been written at least once. See section [0059].

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is

read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

U.S. Pub. No. 2003/0177330 to Idei et al.

Idei et al. relates to a management server for virtual data areas and physical data areas for an array of storage devices. FIG. 1 illustrates a management server 100, storage devices 110, servers 120, network 130 and a special-purpose network 134 for coupling the management server 100 to the respective storage devices 110. See section [0026]. Management server 100 includes access history information 106. FIGS. 2 and 3 illustrate the access history information 106 stored in the memory 108 of the management server 100. The access history information includes read-out history information 300, processor IDS 304 and read-head history information 320. See section [0039]. Read-out history information 300 includes server IDs 302 for identifying a server 120 that issued the instruction, process IDs 304 for indicating the process within the server 120 that issued the instruction 400, virtual volume IDs 306 and time information 312. See section [0040].

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history.

The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

U.S. Pub. No. 2004/0193803 to Mogi et al.

Mogi et al. relates to a cache management method for a storage device. FIG. 1 illustrates a storage system comprising storage devices 40, servers 70 and an administrative server 120. See section [0052]. Each storage device 40 comprises a CUP 12, memory 14 and disk storage units (HDDs) 16. See section [0054]. FIG. 26 illustrates a data structure of HDD performance information 612 including an entry to hold device ID 572, an entry to hold HDD ID 397, and entries to hold performance information 614. See section [0183]. A system management program 140 uses cache monitored statistics information 362, DBMS monitored statistics information 410, and online jobs monitored statistics information 430. System management program 140 also edits these items into a suitable form and stores them in system management information 142 for monitoring history information 510. See section [185].

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as

history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

3ware TwinStor Architecture to 3ware

3ware relates to multiple drives in PCs, servers and workstations for optimizing the maintenance of mirrored data on pairs of ATA disk drives. When data is accessed, TwinStor's system and method employs a profile that it maintains of the disks' layout and an accumulated heuristic history of drive accesses, thus dynamically distributing data retrieval between the drives so that movement of each disk arm is minimized. See Introduction, page 1. A profiling program scans the disk to find the zone breaks, the number of tracks per zone, and other performance information. The result of the profiling is stored in a zone table in a small reserved section on each drive. During execution, the storage switch records an access history to determine whether the current request is best considered a sequential or random access. Separate optimization techniques are applied to these two types of accesses. See Profiling, page 4. For random accesses, a new adaptive algorithm uses the history of previous requests to assign read operations in a way that minimizes the movement of the disk arm. See Adaptive Algorithms, page 5.

As to **claim 1**, the reference does not show or suggest a storage device having a control unit that records data readout locations as a history (i.e., readout access patterns) for a

disk device in a control unit, base on predetermined information. The reference does not show or suggest that data is pre-read from storage based on a command containing information for specifying the history and for specifying the computer accessing the storage.

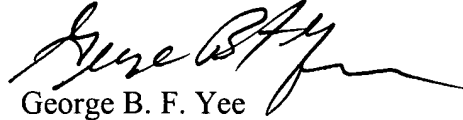
As to **claim 10**, the reference does not show or suggest a storage system having a management computer that transmits a command to a storage device, including information for specifying a history and a computer. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that when the computer reads out data from the storage device, the storage location of that data serves as history information. The reference does not show or suggest that the maintenance computer transmits a second command to the storage device, including information for that specifies a computer and specifies history. The reference does not show or suggest that the storage device reads out data to a cache memory based on the second command.

As to **claim 13**, the reference does not show or suggest a read-ahead method in a storage device. The reference does not show or suggest that a command is transmitted to a storage device, including information for specifying a computer that uses the storage device and a history, or that the location where data is to be stored is recorded as a history when the data is read out by the computer. The reference does not show or suggest that a second command is transmitted, specifying the computer and the history, or that data is pre-read from the storage device, based on the computer and the history specified in the second command.

Conclusion

In view of this comments presented in the instant petition and the claim amendments presented in the accompanying preliminary amendment, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,


George B. F. Yee
Reg. No. 37,478

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
GBFY
60544401 v1



Terry W. Kramer*
 Arlir M. Amado*
 Andreas Baltatzis
 Hans J. Crosby*

Of Counsel
 Tyler S. Brown

Registered Patent Agents
 Thomas A. Powers, Ph.D.
 Matthew J. Gerike
 William W. Lewis

Technology Specialists
 C. Michael Obinna
 Raj C. Patel
 Bijan N. Karimi, M.S.
 Brijesh S. Patel, M.S.
 Paul I. Obinyi
 A. Todd Buttram
 Sung P. Ham, M.S.
 Usha T. Shrestha, M.S., MIP
 Mita Biswas, Ph.D.
 David Groesbeck
 Sung Chan Han, M.S.
 Kyle G. Hepner

*Member Bar other Virginia

April 15, 2005

**PRIVILEGED AND CONFIDENTIAL
 ATTORNEY CLIENT INFORMATION**

Mr. Noboru Otsuka
 HITACHI LTD. INTELLECTUAL PROPERTY GROUP
 IP Development & Management Division
 Patent Department 4
 292, Yoshida-cho, Totsuka-ku
 Yokohama-shi, Kanagawa, Japan 244-0817

RE: Petition-To-Make-Special Search
 For: STORAGE DEVICE HAVING A CAPABILITY
 OF LEARNING ACCESS PATTERNS
 Your Ref. No.: 340301198US01
 Our Ref. No.: HIT 3183

Dear Mr. Otsuka:

We have completed the Petition to Make Special search at the U.S. Patent and Trademark Office regarding the above-identified invention. Enclosed with this letter are our draft Petition to Make Special, and paper and electronic copies of patents and references set forth in our search.

Statement on Prior Art

This search was not provided with an IDS. However, a number of prior art references were discussed in the Background of the Invention section. Our search also included review of the references cited in the Background of the Invention Section and are included in this report if deemed most closely related to the claims. See M.P.E.P. § 708.02, VIII (D).

Search Report

The field of search covered Class 711, subclasses 100, 101, 111, 112, 113 and 114 (U.S. & Foreign). Additionally, a computer database search was conducted on the U.S.P.T.O. systems EAST and WEST for U.S. and foreign patents; a keyword search was conducted in Class 707, and a literature search was also conducted on the internet and commercial databases for relevant non-patent documents. Examiner Stephen Elmer in Class 711 (Group Art Unit 2186) was consulted in confirming the field of search.

Crystal Plaza One
 2001 Jefferson Davis Hwy
 Suite 1101
 Arlington, Virginia 22202
 tel: 703.413.5000
 fax: 703.413.5048

www.kramerip.com

Mr. Noboru Otsuka
April 15, 2005
Page 2

The search was directed towards a storage device having a capability of learning access patterns. In particular, the search was directed towards claims 1, 10 and 13 of U.S. Application No. 10/769,030. With reference to the disclosure, a storage device records data readout locations (i.e. a history) for a disk device to thereby learn access patterns. FIG. 1 illustrates a system having a storage device 101, a management computer 108, a network server 110, and a plurality of computers 111 connected through a network 109. See pg. 6, ln. 23 of the specification. The storage device 101 obtains an access history, such as data readout, from the storage device 101 for each computer 111 and then stores the access history for each computer in the cache control unit 106. See pg. 7 ln. 23-27. FIG. 2 illustrates a configuration of a history information saving table which saves information of the access history to the storage. The table includes a history information management table 301 and a history information saving list 302. See pg. 8, ln. 23-27. With reference to the claims, a storage device records data readout locations (i.e. a history) for a disk device to thereby learn access patterns. Data corresponding to readout locations of the disk device are pre-read and stored in cache memory.

Please note the enclosed documents listed in numerical order for convenience:

U.S. Patent Number

6,834,289 *

6,763,442 *

Inventor(s)

Kaneda et al.

Arakawa et al.

Published Patent Application

2003/0084237

2003/0120674 *

2003/0177330

2004/0193803

Inventor(s)

Yoshida et al.

Morita et al.

Idei et al.

Mogi et al.

Non-Patent Documents

3ware TwinStor Architecture

Author(s)

3ware

* Patent assigned to Hitachi

Brief Description Of The Documents:

U.S. Patent No. 6,834,289 to Kaneda et al.

Kaneda et al. relates to an information processing system and storage area allocating method comprising a plurality of storage devices and a library. FIG. 1 illustrates a system comprising three computers 101, 102, 151, three disk array systems 201, 202, 203, a library system 301 and a fiber channel switch 501. See col. 3, ln. 10-13 and col. 12, ln. 34-43. A cache disk 90 is also provided in computers 101, 102. FIG. 2 illustrates a conversion table 811 stored in computer 151. See col. 5, ln. 51-26. In order to read the data recorded on the medium stored in the library system 301, the computer 151 operates to instruct the library system 301 to convey the medium on which the data to be read is recorded from the shelf to the drive 30. Then, the computer 151 waits for a report on termination of the conveyance from the library system 301 and instructs the drive 30 to read out the data. See col. 4, ln. 30-44.

U.S. Patent No. 6,763,442 to Arakawa et al.

Arakawa et al. relates to a data reallocation system used with multiple data storage systems such as RAIDs. FIG. 1 illustrates a computer and storage system comprising a host 100, disk arrays 200-1, 200-2, a switch 500, clients 800-1, 9800-2 and a local disk 190. Host 100 has a network interface 170 and a Fiber Channel interface 160 and is connected via network 700 to the clients 800-1, 800-2 and disk arrays 200-1, 200-2. See col. 5, ln. 42-54. The local disk 190 stores management information including a logical unit (LU) position name table 191 and an intra-LU address logical position name table 195 to be used by operating system (OS) 120 and file system 110. See col. 6, ln. 8-17. The manager 130 issues a write command of the SCSI standard to the LU for a logical volume and a row of a disk array via the FC interface 160 and writes parameters for information collection as data. See col. 9, ln. 18-22. When the disk array 200-2 receives a read command from the host 100 via FC interface 260-2, the control section 230-2 perceives that the read command concerns the LU and transfers the information readied on the memory 320-2 to the host 100. Control section 300-2 then reports the completion of the read to the host 100. See col. 9, ln. 38-45.

U.S. Pub. No. 2003/0084237 to Yoshida et al.

Yoshida et al. relates to a disk array system controller having a host switch interface section, disk array controlling units, and a cache memory section. FIG. 1 illustrates the disk array controller 1 comprising a plurality of disk array controlling units 1-2 and a host switch interface section 30. See section [0032]. The host switch interface section 30 refers to a management table 31 therein in response to the request

made by the host computer 50 to find an optimum route to the cache memory14. See section [0033]. The host switch interface section 30 is connected through a PATH050, PATH151 and PATHBK 52 with the plurality of channel interface sections of the disk array controlling units. A management table 31 is provided in the host switch interface section 30 in which contains a path selection table 32, a history information table 33, wherein a path selection signal 40 is output to the path selection table 32 on the basis of selected signal PATH 41. See section [0034].

U.S. Pub. No. 2003/0120674 to Morita et al.

Morita et al. relates to a control method for a disk array device and a disk controller wherein a write history management table records histories of conducted writing processes for confirmation of a writing process conducted to a block. FIG. 1 illustrates a disk array device 102, a disk controller 103 and a subgroup of hard disks 114. See section [0054]. A write history management table 113 includes management information for judging whether or not data has been written in the magnetic disk 108. See section [0056]. FIG. 2 illustrates that prepared history flags 301 exist in certain subgroups of hard disks 114 for the M data stripes 201. FIG. 4 illustrates a history flag 301 value of "0" which indicates there is no history of data writing in the respective data stripe, and a flag 301 of "1" which indicates there is a history of data writing in the data stripe, indicating data has been written at least once. See section [0059].

U.S. Pub. No. 2003/0177330 to Idei et al.

Idei et al. relates to a management server for virtual data areas and physical data areas for an array of storage devices. FIG. 1 illustrates a management server 100, storage devices 110, servers 120, network 130 and a special-purpose network 134 for coupling the management server 100 to the respective storage devices 110. See section [0026]. Management server 100 includes access history information 106. FIGS. 2 and 3 illustrate the access history information 106 stored in the memory 108 of the management server 100. The access history information includes read-out history information 300, processor IDS 304 and read-head history information 320. See section [0039]. Read-out history information 300 includes server IDs 302 for identifying a server 120 that issued the instruction, process IDs 304 for indicating the process within the server 120 that issued the instruction 400, virtual volume IDs 306 and time information 312. See section [0040].

U.S. Pub. No. 2004/0193803 to Mogi et al.

Mogi et al. relates to a cache management method for a storage device. FIG. 1 illustrates a storage system comprising storage devices 40, servers 70 and an administrative server 120. See section [0052]. Each storage device 40 comprises a

Mr. Noboru Otsuka
April 15, 2005
Page 5

CUP 12, memory 14 and disk storage units (HDDs) 16. See section [0054]. FIG. 26 illustrates a data structure of HDD performance information 612 including an entry to hold device ID 572, an entry to hold HDD ID 397, and entries to hold performance information 614. See section [0183]. A system management program 140 uses cache monitored statistics information 362, DBMS monitored statistics information 410, and online jobs monitored statistics information 430. System management program 140 also edits these items into a suitable form and stores them in system management information 142 for monitoring history information 510. See section [185].

3ware TwinStor Architecture to 3ware

3ware relates to multiple drives in PCs, servers and workstations for optimizing the maintenance of mirrored data on pairs of ATA disk drives. When data is accessed, TwinStor's system and method employs a profile that it maintains of the disks' layout and an accumulated heuristic history of drive accesses, thus dynamically distributing data retrieval between the drives so that movement of each disk arm is minimized. See Introduction, page 1. A profiling program scans the disk to find the zone breaks, the number of tracks per zone, and other performance information. The result of the profiling is stored in a zone table in a small reserved section on each drive. During execution, the storage switch records an access history to determine whether the current request is best considered a sequential or random access. Separate optimization techniques are applied to these two types of accesses. See Profiling, page 4. For random accesses, a new adaptive algorithm uses the history of previous requests to assign read operations in a way that minimizes the movement of the disk arm. See Adaptive Algorithms, page 5.

While the above-noted Examiner was consulted and confirmed our opinion that the most relevant areas for this invention were reviewed, further searching may uncover additional patents. NOTE: The field of search included the most pertinent areas identified by the Examiner and our office as containing relevant patents.

Enclosed are copies of the cited documents and our invoice for services rendered and disbursements for this matter.

Mr. Noboru Otsuka
April 15, 2005
Page 6

As always, if you have any questions regarding this search, please do not hesitate to call us at (703) 413-5000.

Very truly yours,

Terry W. Kramer
Direct Dial (703) 413-3674
E-mail: terry@kramerip.com

TWK/MJG/TEM/nsa
Enclosure

TABLE OF CONENTS

U.S. Patent No. 6,834,289 to Kaneda et al.	1
U.S. Patent No. 6,763,442 to Arakawa et al.	2
U.S. Patent Application Publication No. 2003/0084237 to Yoshida et al.	3
U.S. Patent Application Publication No. 2003/0120674 to Morita et al.	4
U.S. Patent Application Publication No. 2003/0177330 to Idei et al.	5
U.S. Patent Application Publication No. 2004/0193803 to Mogi et al.	6
Literature 3ware TwinStor TM Architecture by 3ware	7

3ware TwinStor™ Architecture



TwinStor™ Technology: A Compelling Case for Multiple Drives in PCs, Servers and Workstations (August 1999; November 2000; revised April 2002)

Executive Summary

3ware's TwinStor technology provides an optimized method of maintaining mirrored data on pairs of ATA disk drives. Because twin images of the data exist – one image on each drive – backup of valuable data is essentially accomplished each time data is written to the disks.

This safeguard would benefit many of today's computer systems, as most systems contain only a single disk drive that's "protected" by expensive backup hardware and all too often forgotten. With the cost of storage rapidly declining, using a TwinStor-enabled ATA RAID controller, such as 3ware's Escalade 7000 series, in conjunction with multiple ATA disk drives is an inexpensive backup solution that's constantly at work protecting valuable data.

While the inherent fault tolerance of this approach effectively solves backup woes, its prime benefit goes beyond protecting data: an even more compelling aspect of TwinStor is the dramatic performance boost that it also achieves while processing mirrored data. When data is accessed, TwinStor technology employs a profile that it maintains of the disks' layout and an accumulated heuristic history of drive accesses, to dynamically distribute data retrieval between the drives such that movement of each disk arm is minimized – this reduces latency and facilitates streaming. Adaptive algorithms increase performance to the extent that the sequential read bandwidth approaches that of striped (RAID 0) drives and the random transaction rate exceeds that of striped and mirrored (RAID 1) solutions.

A TwinStor-enabled controller plus low-cost ATA drives provide improved performance and fault tolerance over a single-drive configuration and benefit a wide range of applications in home, small office, and server environments.

Introduction

Desktop PCs and small servers are becoming increasingly critical in businesses and homes: The data stored on these systems, from financial records to digital photographs, is often irreplaceable. Disk drive reliability is very high, but the possibility of a drive failure does exist and it is important to make sure that the data remains secure and is not lost. There are many different procedures for backing up data but none are entirely satisfactory. Mirroring the data to a second drive provides an effective and less costly solution than daily back up to secondary media or remote servers.

Consumers will pay premium prices to obtain the highest frequency CPUs but system vendors have typically offered few choices for improving I/O performance (even though many applications are more sensitive to I/O speed than CPU speed). Now that CPU speeds have increased to levels of 1GHz and beyond, this disparity often results in a glaring I/O subsystem bottleneck that hinders application responsiveness. There is however an opportunity to improve the performance of many applications by combining transfer and transaction rates of multiple drives.

The solution that accomplishes this is 3ware's TwinStor technology, which simultaneously provides the fault tolerance of disk mirroring (RAID 1) and the read performance of striping (RAID 0) with superior transaction rates. By using a TwinStor-enabled ATA RAID controller [1], such as 3ware's Escalade 7000, along with low-cost ATA drives, a compelling case can be made for multiple drives per PC.

This white paper discusses trends in data backup strategies and compares the associated costs with the approach offered by TwinStor technology. It also goes under the hood of the technology and uncovers many of the groundbreaking techniques. Within the descriptions of the design concepts, brief tutorials of RAID architecture and disk drive hardware components, offer insights into how TwinStor's algorithms are able to achieve such striking performance gains.

Having laid this groundwork, the discussion concludes with performance and cost comparisons with SCSI that further highlight the benefits of the TwinStor technology. It becomes clear that there are many applications and markets that can effectively make use of TwinStor to prevent data loss and increase storage system efficiency.

Total cost of ownership

The economic justification for mirroring is based on the cost of the second drive compared to the backup and recovery costs after a drive failure. Depending on the business environment, backups can be done several different ways.

Corporate desktop PCs and small servers are often backed up automatically over a network. The total cost of this backup strategy can be quite high when all costs (hardware and software, increased network capacity and system administrators' time) are included. A large corporation recently spent over \$1000 per PC for a centralized hierarchical backup system. Disk mirroring would have solved the daily backup problem at a much lower cost.

Small businesses typically rely on manual backups to tape or other removable media. Total cost of ownership includes risk of losing everything if a backup was not done recently. Even when backups have been done properly, a drive failure may cause any business to close – requiring down time to repair the hardware, reload the operating system and applications and restore the data.

Many home PCs and non-critical business PCs do not adhere to any regular backup procedure, even though the effort to recover from a disk failure could be significant. For these users, the increased performance combined with the added fault tolerance may be a compelling reason to spend a modest amount for a second drive.

Protecting data in this fashion is one of the initial philosophies that spawned RAID technology.

Background of RAID

Mirrored disks have been common in the industry for many years. Disk mirroring, also called shadow sets or RAID 1, uses a pair of disks with the identical data. Every write is sent to both drives to maintain identical copies at all times. Disk mirroring is used in many commercial systems and has been the subject of academic research at the University of California at Berkeley [2] and elsewhere.

Some of the early implementations of mirrored-disk systems attempted to improve random read performance by taking advantage of the separate actuators. The usual technique is to alternate read accesses to the two drives, or to assign reads to drives based on the one that would have the shortest seek time to reach the data. This technique can double the random read performance, but does nothing to improve the streaming read rate because a single drive services each read.

Disk striping of two drives, known as RAID 0, places even blocks of data on one drive and odd blocks on another drive. The main disadvantage of a standard RAID 0 configuration is that reliability is worse than a single drive because a failure of either drive leaves no complete copy of any file.

RAID 5 is another RAID level that provides a way to recover from a drive failure. For each block of data, the parity of N-1 blocks is computed and stored on the nth drive. The primary drawbacks of a RAID 5 configuration is that it requires at least three drives and it sharply decreases the write performance relative to a single drive.

RAID 10, a combination of RAID 1 and RAID 0, provides both data redundancy and improved streaming performance. The drawback of a standard RAID 10 configuration is that it requires four drives but cannot attain more than two times the performance of a single drive.

3ware's TwinStor Technology

TwinStor's mirrored approach optimizes the performance of RAID 1 configurations by algorithmically distributing operations between each drive such that the mechanical overhead of each disk is kept to a minimum. 3ware's new algorithms for intelligent performance optimization, achieve this in several ways:

- Profiling disk drives to obtain drive-specific parameters needed for optimal performance
- Optimizing performance with adaptive algorithms based on the recent access history
- Optimizing for applications that have special performance or reliability requirements

These techniques are applied to pairs of drives, which maintain identical copies of data. All writes are sent to both drives, but reads are free to access whichever copy of the data gives the best performance. Profiling and adaptation are required to carefully orchestrate the actions of both drives to optimize for the best average performance.

To gain insights into how these optimizations are implemented requires a basic understanding of disk drive technology:

A drive contains one or more platters, each with two surfaces and a head per surface. Typical drives today have two to eight heads. All heads are attached to a single actuator, but the fine precision needed to position a head over the data means that the servo electronics can position the actuator to read from only one head at any point in time. Data is organized in tracks; a track contains all the data positioned beneath one head around the entire circumference. Typical disks today have a few hundred 512-byte sectors per track. Outer tracks are longer than inner tracks and hence have more data. Most drives today divide groups of tracks into a small number of zones (16, for instance) and the number of sectors per track stays constant within a zone. Data is typically formatted by starting at the outside of the disk at one head, sequencing through the rest of the heads, and then seeking to the next track location closer to the inside of the disk.

Drives are usually thought of as random access devices. Any drive must seek to position the actuator and wait for the desired data to rotate until it is under the head. Seeks to nearby tracks are much faster than seeks to distant tracks, and seek times can vary from a few milliseconds (ms) to a few tens of ms. The rotational latency depends on the RPM of the drive, with 5,400 RPM, 7,200 RPM and 10,000 RPM drives having maximum latencies of 11.1 ms, 8.3 ms and 6 ms respectively. During sequential accesses within a track, there is no waiting for rotational latency because the needed data is already under the read head. When a sequential access extends beyond a track, a delay of a few ms is required to switch heads. Data is formatted with some skew so the read head is automatically in position to retrieve the next sequential data upon completion of the head switch – this eliminates an entire revolution of the disk that would otherwise be necessary to get to the position of the data. Reading data sequentially can be orders of magnitude faster than reading the data with short random accesses.

The basic idea behind 3ware's TwinStor technology is to reduce seek times and avoid rotational latency by using intelligent algorithms executed by the embedded microprocessor on the disk switch. The high-level flowchart in Figure 1 shows the separate profiling and execution steps.

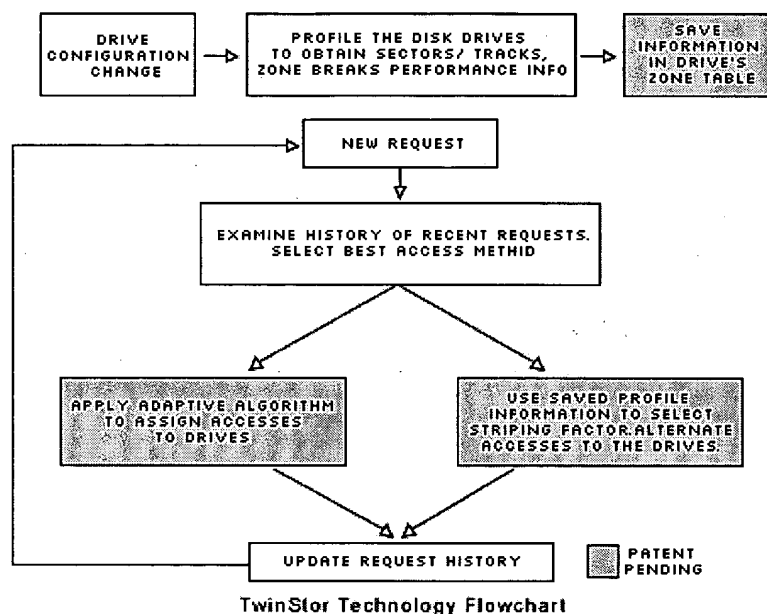


Figure 1. TwinStor Technology Flowchart

Profiling

The first time a new disk is encountered by the storage switch, the profiling program scans the disk to find the zone breaks, the number of tracks per zone, and other performance information. The result of the profiling is stored in a zone table in a small reserved section on each drive. During execution, the storage switch records an access history to determine whether the current request is best considered a sequential or random access. Separate optimization techniques are applied to these two types of accesses.

Adaptive Algorithms

For random accesses, a new adaptive algorithm uses the history of previous requests to assign read operations in a way that minimizes the movement of the disk arm. These optimizations have shown superlinear performance gains on random read operations. Superlinear means that performance gains are better than linear, with two drives giving greater than two times the performance of one drive. In the results shown in Figure 2, the performance gain is about 2.3 times that of a single drive. The reason for this outstanding result is that there are twice as many actuators and each travels less distance than the average distance when only one drive is used.

In most RAID 1 configurations, each I/O is directed to one of the disks and there is no performance improvement if small fixed-length stripes are read alternately from the two drives. For instance, if disk 0 reads the even 32K stripes and disk 1 reads the odd 32K stripes, both disks transfer half the time and spend the other half of the time waiting for the head to pass over data being read by the other drive. This phenomenon is shown on the left side of Figure 2 with small stripe sizes. As the stripe size is increased, it eventually passes the point where the amount of data being skipped is equal to one track.

At this point the data rate increases sharply because there is almost no time wasted for the head to pass over data being transferred by the other drive. At the first peak, the data rate is not quite equal to reading the drive sequentially because one extra disk skew is required when skipping the track read by the other drive. Later peaks have higher bandwidth because the extra skew is spread across more tracks of transferred data. The position of the peaks and the performance at each peak vary depending on the bit density, RPM of the drive and the particular zone being measured.

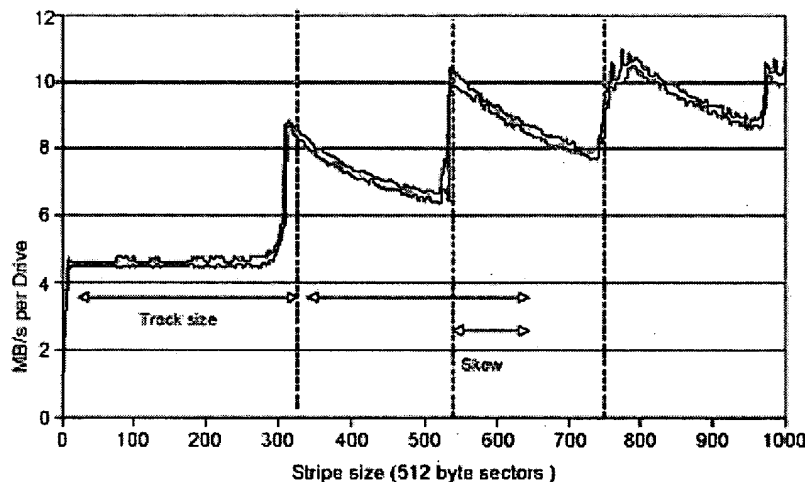


Figure 2. Performance Sensitivity to Stripe Size

The storage switch takes advantage of this phenomenon by setting a stripe size at one of these peaks and simultaneously accessing alternating stripes from the two drives. In this way, long sequential reads run at nearly twice the rate of a single drive. The peaks shift to the left at each zone crossing when moving from the outer diameter of the disk toward the inner tracks. For optimal performance, the zone table is consulted at each zone crossing in order to set the stripe size to the optimal value for that zone. The combination of the sequential and random access methods gives improved performance over a wide range of applications.

Performance

These TwinStor algorithms are implemented within 3ware's Escalade ATA RAID controllers to further enhance its impressive level of throughput. Escalade card's on-board CPU and firmware support this logic and combine it with its packet-switched architecture to achieve breakthrough performance levels from standard, low-cost ATA drives. This results in performance and affordability that can't be attained with competing RAID solutions.

Previously, the highest performing disk architectures were implemented with SCSI drives and RAID controllers from Adaptec, Mylex and others. While SCSI drives were higher in performance than ATA drives, the gap has now closed and many manufacturers use identical head disk assemblies for their ATA and SCSI drives. The market dominance of ATA (87% of the unit volume) assures that the value and availability of high-performance ATA drives will continue to exceed that of SCSI drives.

Figure 3 is a graph comparing a single SCSI drive to a pair of drives using TwinStor technology. Both the SCSI and ATA drives are 7200 RPM. The transfer rate of the ATA drive is slightly higher than the SCSI drive. The access time of the SCSI drive is faster than the ATA drive. The street price of the SCSI drive is more than double that of the ATA drive, making the TwinStor solution less expensive by about 10%. (Escalade controller prices have not been included here, as prices vary among vendors.)

In this comparison, 3ware's TwinStor solutions win in all categories. The streaming read and random read rates are significantly higher than the single SCSI drive. The write rate is slightly higher for the TwinStor solution, showing that the need to write to both disks does not reduce performance relative to writing to a single disk. Not shown is the huge advantage in fault tolerance with the TwinStor solution. Using 9.1 GB drives, a terabyte requires just over 100 drives. With no redundancy and an expected failure rate of one per 500K hours, the SCSI drives would be expected to have about one failure every six months. With the data redundancy in the TwinStor solution, data loss happens only when a second drive fails before there is a chance to repair the first failure. The mean time to failure (MTTF) of the pair is determined based on a mean time to repair (MTTR) of three days using the standard formula: $MTTF_{dual} = MTTF^2 / 2MTTR$ [3]. The chance that the second drive will fail during the three-day repair time is extremely small (over 1,700 years per terabyte).

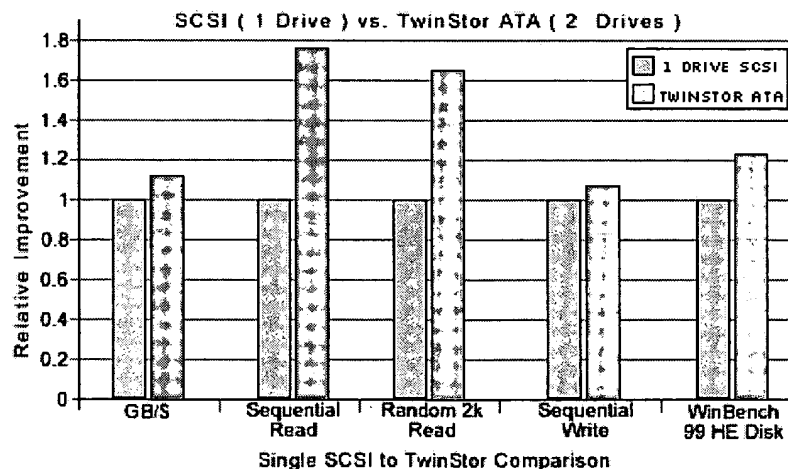


Figure 3. Single SCSI to TwinStor Comparison

Figure 4 shows the comparison between a pair of SCSI drives mirrored in a popular controller with a standard RAID 1

configuration and a pair of ATA drives with TwinStor in an Escalade RAID card. In this comparison, the price advantage over SCSI is even more dramatic. The streaming performance is 73% better, because RAID 1 cannot take advantage of the second drive to improve streaming rates. Random read rates are about the same, even though the SCSI drives used for this test have much faster seek rates than the ATA drives used. The superlinear performance gain obtained from the adaptive random optimization completely makes up for the difference in the drive performance.

Comparison with other classes of SCSI drives shows similar advantages for 3ware's TwinStor technology. For instance, twinned ATA drives compared to a single high-end, 10,000-RPM SCSI drive shows similar overall performance and much better streaming read performance at a much lower cost.

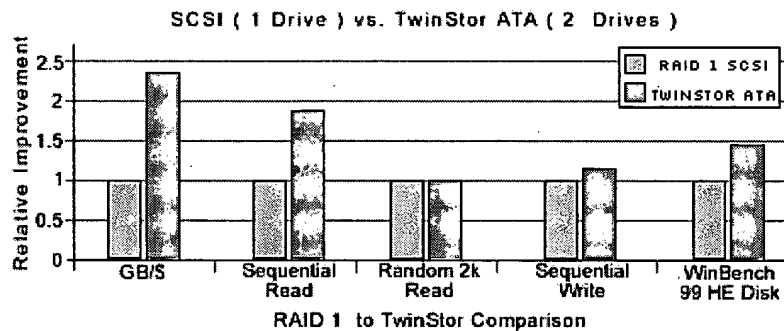


Figure 4. RAID 1 to TwinStor Comparison

Figure 5 shows a four-drive SCSI RAID 5 system (populated with 9.1 GB drives) using an Adaptec RAID controller compared to two TwinStor pair of 18.2 GB ATA drives and the 3ware Escalade card. Each TwinStor pair appears as a single 18.2 GB volume to the NT file system and the two volumes are combined into a single volume with NT software striping. Again, all results favor the TwinStor solution. Capacity is greater because the four-drive RAID 5 solution gives a usable capacity of three times the 9.1 GB drives, while 3ware's TwinStor solution gives twice the capacity of the 18.2 GB drives. Because of the large penalty for writes in RAID 5, the write rates are greatly improved with the TwinStor solution. The WinBench 99 result shows that the overall performance of TwinStor far exceeds the RAID 5 solution and delivers higher capacity with a lower cost.

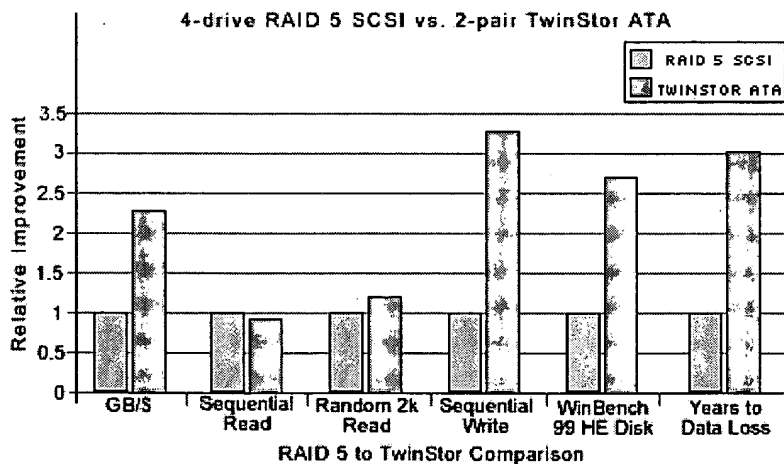


Figure 5. RAID 5 to TwinStor Comparison

Markets

The potential markets for 3ware's TwinStor technology are extremely wide-ranging and should be receptive to improvements in performance and reliability. In the home market where cost is extremely important, the TwinStor technology will be important because many users lack the expertise and tools to back up data regularly or to recover the operating system, applications, and data after a failure. This safeguard, coupled with the ability to "future-proof" the system, may be attractive to a large portion of the home market.

The "SOHO" (small office, home office) market is a natural utilization of TwinStor technology. A disk failure could destroy valuable records and close a small business until the system is recovered. Even if the business never has a disk failure, the peace of mind and increased performance would justify the cost of the second drive and switch.

3ware's TwinStor technology is especially strong in applications which are read intensive and which have a mix of small and large object sizes, the typical transactions performed by a Web server. Systems with the TwinStor technology can be effective in the Web hosting environment and will often show better overall performance than any other way of utilizing a second drive, with the added bonus of fault tolerance.

On-line transaction processing (OLTP) is another area where systems utilizing TwinStor technology will benefit. In the past, OLTP transaction sizes have been very small, with records of just a few hundred bytes, but the trend is definitely towards richer transactions with image, video and audio content. For small, large, or mixed transactions, TwinStor technology provides a good solution at an extremely low cost per transaction. Applications may range from point-of-sale controllers to back-end database systems.

High streaming rates are required in systems that process real-time video and audio. 3ware's TwinStor technology provides this streaming bandwidth and the backup to prevent loss of data when a disk fails. The combination of these requirements makes the TwinStor technology a good fit for a variety of applications.

Conclusion

3ware's TwinStor technology provides a fault-tolerant solution that protects valuable data and improves read performance. This benefits a wide range of computer systems including home PCs, small business systems, on-line transaction point-of-sale controllers, and streaming real-time audio/video processing platforms. The redundancy of data replaces other more costly data backup solutions and eliminates the dilemma that often arises when a drive failure occurs prior to the archival of irreplaceable data.

Given these benefits of data protection and performance enhancement, we are likely to see an increase in the deployment of multiple drive computer systems that make use of TwinStor technology.

References

- [1] "The 3ware StorSwitch Architecture", Technical White Paper, 3ware Inc., April 1999, www.3ware.com. Patterson, D. A.; Gibson, G. A.; and Katz, R. H. "A case for redundant arrays of inexpensive disks (RAID)." Proceedings of the 1988 ACM Conference on Management of Data (SIGMOD). ACM Press, [2] June 1988, 109-116.
- [3] Gray, J. and Siewiorek, D. "High-Availability Computer Systems." Computer, Vol. 24 No. 9, Sept. 1991.

Appendix A. Measurement data

Capacity	27.3	GB	36.4	GB
Street Price	\$1,500	(\$375*4)	\$900	(\$225*4)
GB/\$1000	18.2		40.4	
Sequential Read	52.7	MB/s	47.8	MB/s
Random 2k Read	317	I/Os /sec	379	I/Os /sec
Sequential Write	6.8	MB/s	21.9	MB/s
WinBench HE Disk Test	7,080		19,000	
Years to data loss per TB	587		1,761	
WinBench HE Disk Test	13,300		16,200	
Years to data loss per TB	0.51		1,761	
	SCSI-Seagate		ATA-Quantum	
	Barracuda (RAID 1)		KA (Twin drives)	
Capacity	9.1	GB	9.1	GB
Street Price	\$750	(\$375*2)	\$336	(\$168*2)
GB/\$1000	12.1		27.1	
Sequential Read	18.6	MB/s	32.1	MB/s
Random 2k Read	201	I/Os /sec	200	I/Os /sec
Sequential Write	18.4	MB/s	19.3	MB/s
WinBench HE Disk Test	13,300		16,200	
Years to data loss per TB	1,761		1,761	
	SCSI-Seagate		ATA-Quantum	
	Cheetah(JBOD)		KA (Twin drives)	
Capacity	9.1	GB	9.1	GB
Street Price	\$514		\$336	(\$168*2)
GB/\$1000	17.7		27.1	
Sequential Read	26.1	MB/s	32.1	MB/s
Random 2k Read	155	I/Os /sec	200	I/Os /sec
Sequential Write	23.9	MB/s	19.3	MB/s
WinBench HE Disk Test	16,400		16,200	
Years to data loss per TB	1		1,761	
	SCSI-Seagate		ATA-Quantum	

	Barracuda(RAID 5)		KA(2 pairs of Twins)	
Capacity	27.3	GB	36.4	GB
Street Price	\$1,500	(\$375*4)	\$900	(\$225*4)
GB/\$1000	18.2		40.4	
Sequential Read	52.7	MB/s	47.8	MB/s
Random 2k Read	317	I/Os/sec	379	I/Os/sec
Sequential Write	6.8	MB/s	21.9	MB/s
WinBench HE Disk Test	7,080		19,000	
Years to data loss per TB	587		1,761	

Appendix B. Test conditions

System Configuration

500MHz - Single Processor Pentium III - 128Meg SDRAM
Windows NT 4.0 - w/Service Pack 5
ATI Rage Pro Turbo w/ 8Meg- 1024x768, 64K Colors
NEC 40X IDE CD-ROM

ATA	Quantum Fireball Plus KA 9.1 G and 18.2 G, 7200 RPM
	3ware Disk Switch 4 Controller
SCSI	Seagate ST39175LW Hard Drive 9.1G, 7200 RPM
	Seagate Cheetah Hard Drive 9.1G, 10,000 RPM
	Adaptec AAA-131U2 PCI RAID Controller

Prices from www.dirtcheapdrives.com 8/14/99

WinBench 99 tests on 2 GB Fat partition
Iometer tests on Full Volume NTFS

Reliability Assumptions

7,200 RPM drives (Quantum and Seagate): 500K Hour MTBF per drive
10,000 RPM drives (Cheetah): 1M hour MTBF
Three-day repair time

BEST AVAILABLE COPY